

# 3차 의존관계에 기반한 곱 근사를 이용한 다수 인식기의 결합

(Combining Multiple Classifiers using Product Approximation  
based on Third-order Dependency)

강 희 중 <sup>†</sup>

(Hee-Joong Kang)

**요약** 인식기와 클래스 레이블로 구성된 고차 확률 분포를 가정이나 근사 없이 저장하고, 평가하는 것은 기하급수적으로 복잡하고 관리하기 어렵다. 따라서, 의존관계를 이용한 근사 방법에 의존하게 된다. 본 논문에서는 기존의 2차 의존관계에 기반한 곱 근사 방법을 확장하여, 이 확률 분포를 3차 의존관계에 의해 최적으로 곱 근사하는 방법을 제안하고자 한다. 제안된 3차 의존관계에 기반한 곱 근사 방법은 Concordia 대학과 UCI(University of California, Irvine) 대학으로부터 얻은 필기 숫자를 인식하는 실험에서 다수 인식기의 결합 방법에 적용되었고, 실험을 통하여 제안된 방법의 유용성을 살펴보았다.

**키워드** : 3차 의존관계, 다수 인식기의 결합, 곱 근사, 유사도, 베이지안 결합 방법, 필기 숫자

**Abstract** Storing and estimating the high order probability distribution of classifiers and class labels is exponentially complex and unmanageable without an assumption or an approximation, so we rely on an approximation scheme using the dependency. In this paper, as an extended study of the second-order dependency-based approximation, the probability distribution is optimally approximated by the third-order dependency. The proposed third-order dependency-based approximation is applied to the combination of multiple classifiers recognizing handwritten numerals from Concordia University and the University of California, Irvine and its usefulness is demonstrated through the experiments.

**Key words** : third-order dependency, combining multiple classifiers, product approximation, measure of closeness, Bayesian combination method, handwritten numeral

## 1. 서론

다수 인식기의 결합 방법은 인식기의 인식 결과 유형인 단일 선택, 측정 점수 리스트, 순위 리스트 등에 따라 각각 유형-I, 유형-II, 유형-III으로 구분할 수 있다 [1,2]. 유형-I 결합 방법은 결합되는 인식기의 인식 결과가 단 하나의 클래스로만 출력되는 단일 선택을 가정하여 다수 인식기를 결합하는 방법이고, 유형-II는 인식 결과가 클래스들의 순위로 출력된다고 가정하여 결합하며, 유형-III은 클래스들의 인식 점수로 출력된다고 가정하여 결합한다. 이 가운데 유형-I 결합 방법은 인식기의 인식 결과 유형에 덜 의존적이고, 임의의 인식기도 결합할 수 있다는 점에서 유용하다. 이러한 유형-I 결합

방법에는 투표 방법[1], Behavior-Knowledge Space (BKS) 방법[3], 증거적 추론에서 사용되는 Dempster-Shafer 방법[4], 독립 가정을 취하는 베이지안 방법[1,5], 1차 및 2차 의존관계에 기반한 곱 근사를 취하는 베이지안 방법[2] 등이 있다. 본 논문에서는 기존의 2차 의존관계에 기반한 곱 근사를 취하는 베이지안 방법을 고려하여, 이를 3차 의존관계에 까지 확장하려 한다.

확률 이론과 베이지안 방법을 이용하여 유형-I 결합 방법을 수식으로 표현하면 다음과 같다. 입력  $x$ 가  $K$ 개의 인식기 (예를 들면,  $E_1, \dots, E_K$ )에 병렬로 주어질 때, 그들의 인식 결과는 다음과 같은  $K$ -차원의 결정 벡터인  $C = \langle E_1(x) = M_1, \dots, E_K(x) = M_K \rangle$ 가 관찰된다. 여기서,  $L$ 개의 클래스 레이블(label)은  $M = \{M_1, \dots, M_L\}$ 으로 표현된다. 개별 인식기는 주어진 입력에 대해 이들 클래스 중의 하나를 선택한다. 이러한  $K$ -차원의 결정 벡터와 이에 대해 참(true)인 클래스 레이블 집합은 BKS 방법에서는 고차 빈도수 테이블을 만드는데 사용

· 위 연구는 한국과학재단의 박사후연수지원사업으로부터 지원을 받았음

<sup>†</sup> 종신회원 : 한성대학교 컴퓨터공학부 교수

hjkang@hansung.ac.kr

논문접수 : 2003년 10월 22일

심사완료 : 2004년 3월 13일

되고, 베이지안 방법에서는  $(K+1)$ -차원의 이산 확률 분포를 구성하고 평가하는데 사용된다.

베이지안 방법을 이용하여 다수 인식기를 결합하는 방법은 주어진  $K$ 개의 인식 결과에 대해서 개별 클래스 별로 사후(posterior) 확률  $P$ 를 계산하고, 그 사후 확률  $P^*$ 가 최대가 되는 가설 클래스  $m$ 을 결정하는 것이다. 즉, 클래스 집합이  $M$ 일 때, 최대 사후 확률  $P^*$ 는 다음과 같다:  $\max_m P(m \in ME_1(x) = M_1, \dots, E_K = M_K)$ . 이러한 사후 확률의 계산을 위하여 훈련 단계에서 샘플 데이터로부터  $(K+1)$ -차원의 이산 확률 분포가 구성되어야 하고, 다음과 같은  $(K+1)$ -차원의 확률 분포를 계산할 수 있어야 한다:  $P(m, E_1(x) = M_1, \dots, E_K = M_K)$ .

그러나, 이와 같은 고차 확률 분포를 독립 가정이 없이 다루는 것은  $K$ 가 커질 때 BKS 방법과 같이 이론적으로 기하급수적인 복잡도를 지니게 되어, 의존관계를 이용한 곱 근사 방법에 의존하게 된다. Chow와 Liu는  $n$ -차원의 확률 분포를 1차 의존관계에 의한  $(n-1)$ 개의 2차 성분 분포의 곱으로 최적으로 근사하는 방법을 제안하였다[6]. 강희중 등은 이러한 의존관계의 개념을 다수 인식기의 결합에 적용하여, Chow와 Liu가 사용한 유사도(measure of closeness)를 이용하여 2차 의존관계에 의한 최적의 곱 근사 집합을 구하는 새로운 방법과 관련 프레임워크를 제안하였다[2].

본 논문에서는 2차 의존관계에 의한 방법의 확장으로서, 확률 분포를 3차 의존관계에 의하여 최적으로 곱 근사하는 새로운 근사 방법을 제안하고, 이러한 근사 방법을 이용하여 다수 인식기를 결합하고자 한다. 3차 의존관계는 저차 의존관계 보다 근사 오류를 더 줄일 수 있으므로, 대표성을 지닌 훈련 데이터가 가용한 경우에 저차 의존관계에 의한 인식을 보다 더 높은 인식을 잠재적으로 제공할 수 있으리라 기대된다. 새로이 제안된 결합 방법은 Concordia 대학의 CENPARMI((Center for Pattern Recognition and Machine Intelligence) 연구소와 UCI(University of California, Irvine) 대학으로부터 얻은 무제한 필기 숫자를 인식하는 인식기들로 평가되었다.

본 논문의 구성은 다음과 같다. 2장에서는 다수 인식기를 결합하는데 있어서 의존관계에 기반한 곱 근사 방법에 대해서 소개하고, 3장에서는 3차 의존관계에 기반한 확률 분포를 최적으로 곱 근사하는 방법과 이를 이용한 베이지안 결합 방법을 설명한다. 4장에서는 제안된 방법을 다수 인식기의 결합에 적용한 실험 결과에 대해서 기술하고, 5장에서 결론을 맺는다.

## 2. 다수 인식기의 결합에서 의존관계에 기반한 곱 근사 방법

인식기가 서로 독립적으로 수행한다는 가정은 인식을 향상이라는 측면에서 인식기들이 다른 인식기에 통계적으로 의존적인 성향을 보이기 때문에 확고한 것은 아니라고 본다. 따라서, 다수의 인식기를 결합하는데 있어서 인식기들 간의 의존관계를 고려하는 것이 바람직하다. 의존관계에 기반하여 다수의 인식기를 결합하는 것은 다음의 순차적인 두 단계로 구성된다: 첫째는 의존관계에 의하여 근사하는 단계이고, 둘째는 근사 결과를 이용하여 베이지안 방법으로 확률적인 결합을 수행하는 단계이다. 의존관계에 의하여 근사하는 단계는 주어진 의존관계의 차수( $d$ , 단  $1 \leq d \leq K$ )에 따라 주어진  $(K+1)$ -차원의 확률 분포에 대한 최적의 곱 근사를 구하는 것이다. 베이지안 방법을 사용하여 확률적인 결합을 수행하는 단계는 이전의 근사 단계에서 구한 최적의 곱 근사를 베이지안 결정 규칙에 적용하여 실질적으로 다수 인식기의 결과를 결합하는 단계이다.

본 논문에서는 이러한 의존관계가 평균 상호정보를 계산함으로써 통계적으로 측정될 수 있다고 가정한다. 평균 상호정보는 한 인식 결과인 특별한 사건의 발생이 다른 대안의 발생 가능성에 대해서 얼마나 많이 나타낼 수 있는가에 관한 정량적 측정치로서 정의되어 있다[6]. 이러한 평균 상호정보에 관한 수학적 근거는 Lewis가 제안한 유사도(measure of closeness)의 정의로부터 유도된다[7]. 따라서, 의존관계는  $(K+1)$ -차원의 확률 분포를 저차 성분 분포의 곱으로 근사하기 위한 이론적 근거를 제공하게 된다.

의존관계에 기반한 근사 방법은 여러 측면에서 독립 가정에 기반한 베이지안 방법과 BKS 방법 사이에서 중간적인 역할을 한다.  $d$ 차 의존관계를 고려하는 것은 이를 이용한 베이지안 방법에서의 필요 저장 공간을  $(K+1-d) \cdot L^{d+1}$  만큼 요구하게 되고, BKS 방법에서와 같은 잠재적인 높은 기각률을 차원 축소에 의하여 낮추도록 한다. 즉, 필요 저장 공간의 복잡도인  $O(L^{d+1})$ 는 독립 가정에 기반한 베이지안 방법에서의 복잡도인  $O(L^2)$ 와 BKS 방법에서의 복잡도인  $O(L^{K+1})$  사이에 놓이게 된다. 예를 들어, 3차 의존관계를 고려한 경우의 복잡도는  $O(L^4)$ 이다. 의존관계의 차수  $d$ 는 더 나은 근사나 성능을 위해서 자원이 허용되는 한 조정될 수 있다. 그러나, 가장 높은 성능을 얻기 위하여 의존관계의 차수  $d$ 를 어떻게 적절히 선정해야 하는가에 관해서는 여전히 중요한 연구 주제로 남아 있다.

$K$ 개의 인식기를 결합하기 위하여,  $d$ 차 의존관계에 의한 최적의 곱이 근사 단계에서 구해져야 한다. 변수  $V$ 를 클래스 레이블과  $K$ 개의 인식 결과로 이루어진  $(K+1)$ -차원의 변수라고 정의하자.  $(K+1)$ -차원 확률 분

포가  $d$ 차 의존관계에 의해 근사될 때, 근사 분포  $P_a(V)$ 가 실제 분포인  $P(V)$ 에 얼마나 가까운가를 측정하기 위한 기준이 필요하다. 정보 이론에 바탕을 둔 그러한 기준은 Lewis에 의해 제안되었으며, 유사도라고 불린다. 유사도  $I$ 는 수학적으로 다음과 같이 정의된다:

$$I(P(V), P_a(V)) = \sum P(V) \log \frac{P(V)}{P_a(V)} \quad (1)$$

이와 같이, 유사도는 실제 확률 분포인  $P(V)$ 에 포함된 정보와 근사 확률 분포인  $P_a(V)$ 에 포함된 정보와의 차이로서 정의되며, 그 차이가 작을수록 이들 간의 유사성은 높다고 할 수 있다. 따라서, 최적의 근사 확률 분포는 유사도  $I$ 의 값을 최소화시키는 것에 의해서 구해질 수 있다.

### 3. 3차 의존관계에 기반한 곱 근사 방법

이 장에서는 3차 의존관계를 기반으로  $(K+1)$ -차원 확률 분포를 최적으로 곱 근사하는 방법을 자세히 기술한다. 수식의 간단한 표기를 위하여,  $(K+1)$ -차원 확률 분포에서  $E_j(x) = M_j$ 를  $C_j$ 으로,  $m \in M$ 를  $C_{K+1}$ 로 각각 표기한다. 즉,  $(K+1)$ -차원 확률 분포  $P(V)$ 는 또한  $P(C_1, \dots, C_{K+1})$ 으로 표현된다. 고차 확률 분포의 곱 근사를 구하는 방법에 있어서 1차 의존관계와 2차 의존관계를 고려하는 것은 참고문헌 [2]에 기술되어 있으며, 이제부터는 3차 의존관계에 의해 최적의 곱 근사를 구하는 새로운 방법에 대해서 기술한다.

3차 의존관계를 고려할 때, 예상되는 근사 확률 분포는 아래와 같이 4차 구성 분포의 곱으로 정의될 수 있다:

$$P_a(C_1, \dots, C_{K+1}) = \prod_{j=1}^{K+1} P(C_n | C_{n_{a(j)}}, C_{n_{a(j)}}, C_{n_{a(j)}}), \quad (2)$$

$$(0 \leq \beta(j), \alpha(j), \alpha(j) < j)$$

여기서  $C_n$ 은  $C_{n_{a(j)}}, C_{n_{a(j)}}, C_{n_{a(j)}}$ 에 모두 조건부로 정의되고,  $(n_1, \dots, n_{K+1}), (n_{\beta(1)}, \dots, n_{\beta(K+1)}), (n_{\alpha(1)}, \dots, n_{\alpha(K+1)})$ , 또는  $(n_{a(1)}, \dots, n_{a(K+1)})$ 은 정수  $(1, \dots, K+1)$ 의 미지의 순열로 위의 식 (2)에 있는 근사 확률 분포 정의식의 제한조건을 만족시켜야 하며,  $C_0$ 은 널(null) 성분이다. 그리고, 정의에 의하여  $P(C_n | C_0, C_0, C_0)$ 은  $P(C_n)$ 이고,  $P(C_n | C_0, C_0, C_{n_{i(j)}})$ 은  $P(C_n | C_{n_{i(j)}})$ 이고,  $P(C_n | C_0, C_{n_{i(j)}}, C_{n_{i(j)}})$ 은  $P(C_n | C_{n_{i(j)}}, C_{n_{i(j)}})$ 이며, 여기서  $C_{n_{i(j)}}$ 은  $C_{n_{a(j)}}, C_{n_{a(j)}},$  또는  $C_{n_{a(j)}}$ 을 의미한다. 위의  $(K+1)$ -차원 확률 분포  $P$ 에 3차 의존관계에 의한 최적의 곱 근사를 위하여 제안된 다음의 알고리즘을 적용하면, 구해진 최적의 곱 근사 집합으로부터 미지의 순열  $(n_1, \dots, n_{K+1})$ 와 이에 대한 조건부 순열들인

$(n_{\beta(1)}, \dots, n_{\beta(K+1)}), (n_{\alpha(1)}, \dots, n_{\alpha(K+1)})$ , 그리고  $(n_{a(1)}, \dots, n_{a(K+1)})$ 의 값을 결정할 수 있게 된다.

한편, 위의 식 (2)에서  $C_{n_{a(j)}}$ 가 모든  $C_n$ 에 동일하다면, 즉  $C_n$ 가 주어진  $C_{K+1}$ 에 대하여  $C_{n_{a(j)}}$ 와  $C_{n_{a(j)}}$ 에 조건부로 의존한다고 가정하면, 근사 확률 분포는 아래와 같은 4차 구성 분포의 곱으로 정의될 수 있다:

$$P_a(C_1, \dots, C_{K+1}) = \prod_{j=1}^{K+1} P(C_n | C_{n_{a(j)}}, C_{n_{a(j)}}, C_{K+1}), \quad (3)$$

$$(0 \leq \beta(j), \alpha(j) < j)$$

이러한 근사 분포는 3차 의존관계에 의한 곱 근사의 특수한 경우로서 조건부 2차 의존관계에 의한 근사 분포라고 정의할 수 있다.

이제부터는 식 (1)로 정의된 유사도  $I$ 에 식 (2)의 3차 의존관계에 의한 근사 분포를 적용하여 어떻게 최적의 곱 근사 분포를 구할 수 있는가에 관한 수학적 근거를 살펴보기로 한다. 간단한 수식 표현을 위해서, 인식기 변수의 아래 첨자  $n$ 을 생략하여  $C_n$ 를  $C_j$ 으로 표기한다.

$$I(P(V), P_a(V)) = \sum P(V) \log \frac{P(V)}{P_a(V)}$$

$$= \sum P(V) \log P(V) - \sum_{j=1}^{K+1} \sum P(V) \times \log P(C_j | C_{\beta(j)}, C_{\alpha(j)}, C_{a(j)})$$

$$= - \sum_{j=1}^{K+1} U(C_j; C_{\beta(j)}, C_{\alpha(j)}, C_{a(j)}) + \sum_{j=1}^{K+1} H(C_j) - H(V) \quad (4)$$

$$H(V) = - \sum P(V) \log P(V) \quad (5)$$

$$H(C_j) = - \sum P(V) \log P(C_j) = - \sum P(C_j) \log P(C_j) \quad (6)$$

$$U(C_j; C_{\beta(j)}, C_{\alpha(j)}, C_{a(j)}) = \sum P(V) \log \frac{P(C_j | C_{\beta(j)}, C_{\alpha(j)}, C_{a(j)})}{P(C_j)}$$

$$= \sum P(C_j; C_{\beta(j)}, C_{\alpha(j)}, C_{a(j)}) \times \log \frac{P(C_j | C_{\beta(j)}, C_{\alpha(j)}, C_{a(j)})}{P(C_j)} \quad (7)$$

위의 식 (4)로부터, 최적의 곱 근사 분포인 근사 확률 분포  $P_a$ 를 구하기 위하여 유사도  $I$ 를 최소화하는 것은

$H(V)$ 와  $\sum_{j=1}^{K+1} H(C_j)$ 인 엔트로피 항이 상수이기 때문에,

평균 3차 상호정보의 합인  $\sum_{j=1}^{K+1} U(C_j; C_{\beta(j)}, C_{\alpha(j)}, C_{a(j)})$ 를 최대화하는 것이다. 따라서, 제안되는 알고리즘은 가능한 모든 곱 근사 집합으로부터 근사 분포의 정의식에 명시된 제한조건을 만족하면서  $\sum_{j=1}^{K+1} U(C_j; C_{\beta(j)}, C_{\alpha(j)}, C_{a(j)})$ 이 최대가 되는 3차 의존관계의 최적의 곱 근사 집합을 찾는 것이다. 이러한 알고리즘을 간략히 작성하

면 아래와 같다.

입력 :  $K$ 개의 인식 결과와 클래스 레이블로 구성된  $w$  개의 샘플 데이터  $S^1, \dots, S^w$

출력 : 평균 상호정보에 대한 3차 의존관계에 의한 최적의 곱 집합

- 방법 : 1. 샘플로부터 2차, 3차, 4차 구성 분포를 계산, 평가한다.  
 2. 평가된 구성 분포로부터 1차, 2차, 3차 평균 상호정보를 계산한다.  
 3. 계산된 평균 상호정보로부터 1차, 2차, 3차 의존관계에 의한 구성 분포의 평균 상호정보의 최대 합을 구하고, 관련 곱 근사 분포를 결정한다.

maxTweight=0;

for n1=1 to 1차 의존관계 분포의 수 (즉,  $\frac{(K+1)K}{2}$ ) do

T1 = 선정된 1차 의존관계 분포의 평균 상호정보;

for n2=1 to 2차 의존관계 분포의 수 (즉,  $\frac{(K+1)K(K-1)}{2}$ ) do

T2 = 선정된 1차 의존관계에 따라 선정된 2차 의존관계 분포의 평균 상호정보;

T3 = 0;

for n3=1 to 탐색되지 않은 인식기의 수 do

탐색되지 않은 인식기중의 하나를 선정;

선정된 인식기와 관련된 가장 큰 허용 가능한 3차 의존관계에 의한 구성 분포를 선정;

T3 = MAX(T3, T1+T2+(선정된 3차 의존관계 분포의 평균 상호정보));

end

end

maxTweight = MAX(maxTweight, T3);

maxTweight와 관련된 1차, 2차, 3차 의존관계에 의한 구성 분포를 저장;

end

최대의 maxTweight와 관련되어 저장된 1차, 2차, 3차 의존관계에 의한 근사 분포의 곱 집합 구하기;

위 알고리즘은 의존관계의 차수가 1에서 3까지 증가 하면서, for 반복문을 통하여 허용 가능한 곱 근사 집합을 구성하면서 평균 상호정보의 합을 누적하고, max-Tweight가 최대가 되는 곱 근사 집합을 선정한다. 제안된 알고리즘의 계산 복잡도는  $O(n^3)$ 으로, 훈련 샘플의 개수가 많아질수록 큰 문제가 될 수 있다. 따라서, 향후 이 복잡도를 줄이기 위한 노력이 필요하다고 본다. 그렇지만, 이 알고리즘은 훈련 샘플로부터 단지 한 번만 수행되기 때문에, 작은 규모의 훈련 샘플에 대해서는 이러한 계산 복잡도가 제안된 방법을 다수 인식기의 결합에 적용하는데 중요하지는 않다. 한편, 2차 의존관계를 고려하여 최적의 곱 근사를 구하는 알고리즘의 계산 복잡도는  $O(n^2)$ 이고, 1차 의존관계의 경우에는 최대 가중치

알고리즘을 사용하기 때문에 계산 복잡도가  $O(n \log n)$ 이다.

이와 같이 3차 의존관계에 의한 최적의 곱 근사 집합으로부터 구해진 근사 확률 분포는 실제로 다수 인식기를 결합하기 위하여 베이저안 방법에 적용된다.  $K$ 개 인식기의 베이저안 방법에 의한 결합은 베이저안 정리의 근사 확률 분포로부터 유도된다. 가설 클래스  $m$ 에 대하여 지지되는 신뢰 함수  $Bel(m)$ 은 다음과 같이 클래스  $m$ 의 사후 확률로서 정의된다:

$$Bel(m) = P(m \in M | C_1, \dots, C_K) \quad (8)$$

베이저안 정리의 3차 의존관계의 최적의 곱 근사를 이용하고, 클래스를 나타내는  $m \in M$ 을  $C_{K+1}$ 으로 표현함으로써, 식 (2)와 (8)로부터 아래와 같은 신뢰 함수식을 정의하게 된다:

$$\begin{aligned} Bel(m) &= P(m \in M | C_1, \dots, C_K) \\ &= \frac{P(C_1, \dots, C_K, C_{K+1})}{P(C_1, \dots, C_K)} \\ &= \frac{\sum_{j=1}^{K+1} P(C_n | C_{n_{a(j)}}, C_{n_{b(j)}}, C_{n_{d(j)}})}{P(C_1, \dots, C_K)} \\ &\approx \eta \sum_{j=1}^{K+1} P(C_n | C_{n_{a(j)}}, C_{n_{b(j)}}, C_{n_{d(j)}}) \end{aligned} \quad (9)$$

여기서  $\eta = \sum_{j=1}^{K+1} Bel(M_j) = 1$ 이기 위한 상수이고,  $(n_1, \dots, n_{K+1})$ ,  $(n_{b(1)}, \dots, n_{b(K+1)})$ ,  $(n_{d(1)}, \dots, n_{d(K+1)})$ , 또는  $(n_{a(1)}, \dots, n_{a(K+1)})$ 은 정수  $(1, \dots, K+1)$ 의 미지의 순열이다. 주어진 결정 벡터  $C$ 로부터 계산된 신뢰 함수  $Bel(m)$ 에 준하여, 최대화된 사후 확률  $P^*(M_j \in M | C_1, \dots, C_K)$ 을 선택할 수 있고, 아래와 같은 결정 규칙  $D(C)$ 에 따라서 결합된 인식 결과가 결정되거나 또는 기각된다:

$$D(C) = \begin{cases} M_j, & \text{if } Bel(M_j) = \max_{M_i \in M} Bel(M_i) \\ L+1, & \text{otherwise.} \end{cases} \quad (10)$$

이러한 체계적인 의존관계 기반의 곱 근사 방법을 사용함으로써, 자원이 허용하는 한도 내에서 고려할 의존관계의 차수는 더 증가될 수 있다.  $d$ 차 의존관계에 의한 최적의 곱 근사 집합은 아래의 근사 확률 분포와 같이, 하나의 1차 의존관계, 하나의 2차 의존관계, ..., 하나의  $(d-1)$ 차 의존관계, 그리고  $(K-d)$ 개의  $d$ 차 의존관계에 의한 구성 분포들로 구성된다:

$$P_d(C_1, \dots, C_{K+1}) = \prod_{j=1}^{K+1} P(C_n | C_{n_{a(j)}}, \dots, C_{n_{a(d)}}), \quad (11)$$

$$(0 \leq id(j), \dots, d(j) < j)$$

여기서  $C_n$ 은  $C_{n_{a(d)}}$ 부터  $C_{n_{a(1)}}$ 까지 허용할 수 있는 최대  $d$ 개의 성분에 조건부로 정의되고,  $(n_1, \dots, n_{K+1})$ ,  $(n_{a(1)}, \dots, n_{a(K+1)})$ , ..., 또는  $(n_{d(1)}, \dots, n_{d(K+1)})$ 은 정수  $(1, \dots, K+1)$ 의 미지의 순열로 위의 식 (11)에 있는 근사 확률 분포 정의식의 제한조건을 만족시켜야 하며,

$C_0$ 은 널(null) 성분이다. 식 (11)은 ( $d=K$ )일 때, 확률의 체인 규칙의 정의와 같은 것처럼 보인다. 왜냐하면, 그 체인 규칙은  $K$ 차 의존관계에 의한 곱 근사의 하나이기 때문이다.

식 (11)에서 보여진 대로 3차 의존관계 이상의  $d$ 차 의존관계가 인식기들의 베이저안 결합 방법에 고려되면,  $Bel(m)$ 은 식 (8)과 (11)로부터 아래와 같은 신뢰 함수 식으로 정의된다:

$$\begin{aligned}
 Bel(m) &= P(m \in MC_1, \dots, C_K) \\
 &= \frac{P(C_1, \dots, C_K, C_{K+1})}{P(C_1, \dots, C_K)} \\
 &= \sum_{j=1}^{K+1} \frac{P(C_j) C_{n_{id(j)}, \dots, C_{n_{a(d)}}}{P(C_1, \dots, C_K)} \\
 &\approx \eta \sum_{j=1}^{K+1} P(C_j) C_{n_{id(j)}, \dots, C_{n_{a(d)}}
 \end{aligned} \tag{12}$$

여기서  $\eta = \sum_{j=1}^{K+1} Bel(M_j) = 1$ 이 위한 상수이고, ( $n_1, \dots, n_{K+1}$ ), ( $n_{id(1)}, \dots, n_{id(K+1)}$ ), ..., 또는 ( $n_{a(1)}, \dots, n_{a(K+1)}$ )은 정수 ( $1, \dots, K+1$ )의 미지의 순열이다. 이러한  $d$ 차 의존관계에 기반한 다수 인식기의 결합을 위해서 식 (12)에 정의된  $Bel(m)$ 을 식 (10)의 결정 규칙  $D(C)$ 에 적용할 수 있다.

#### 4. 다수 인식기의 결합 실험 결과 및 분석

제안된 3차 의존관계 기반의 곱 근사 방법을 실제 다수 인식기의 결합 방법에 적용하기 위하여 6개의 숫자 인식기 (즉, E1, E2, E3, E4, E5, E6)를 실험에 사용하였다. 이들 인식기는 Matsui 등이 제안한 칸투어(contour) 특징[8]과 Oh와 Suen이 제안한 메쉬(mesh), 픽셀 거리 함수(pixel distance function), 방향성 거리 분포(directional distance distribution) 특징[9] 등을 사용하거나, Oh 등이 제안한 숫자의 구조적 정보, 예를 들면, 바운딩 박스(bounding box), 중심(centroid), 수평선(run) 또는 획(stroke)의 넓이와 같은 특징[10] 등을 사용하여 KAIST와 전북대학교에서 개발되었다. 이들 인식기에 대한 개략적인 특성은 표 1과 같다. 특히, 인식기 E4와 E5는 Concordia 대학의 숫자 데이터로부터 추출된 구조적 지식에 의해서 개발되었기 때문에, UCI

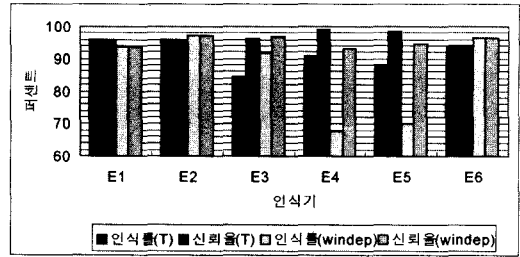


그림 1 개별 인식기의 인식 성능

대학의 숫자 데이터에 대해서는 그리 좋은 편이 아니다. 테스트 데이터 세트  $T$ 와  $windep$ 에 대한 개별 인식기의 인식 성능은 그림 1과 같다.

인식 실험에 사용된 필기 숫자 데이터베이스를 간략히 설명하면 다음과 같다. 첫째, Concordia 대학 CENPARMI 연구소의 데이터 세트는 크게 두 개의 훈련 데이터 세트인  $A$ ,  $B$ 와 한 개의 테스트 데이터 세트인  $T$ 로 이루어져 있다[11]. 이들 각 데이터 세트는 클래스당 200개의 샘플을 가지고 있다. 둘째, UCI (University of California, Irvine) 대학의 광학 인식을 위한 데이터 세트는 크게 두 개의 훈련 데이터 세트인  $tra$ ,  $cv$ 와 두 개의 테스트 데이터 세트인  $wdep$ ,  $windep$ 로 이루어져 있다[12].  $tra$  데이터 세트는 클래스당 약 190개의 샘플을 가지고 있고,  $cv$ 와  $wdep$  데이터 세트는 약 95개의 샘플을,  $windep$  데이터 세트는 약 180개의 샘플을 가지고 있다.  $windep$  데이터 세트는 훈련 데이터의 필기자와 무관한 데이터이다. 신경망 기반의 인식기는 훈련 데이터 세트인  $A$ 와  $tra$ 로 각각 훈련되었다. 의존관계 기반의 곱 근사를 구하기 위하여 각각 두 개의 데이터 세트인  $A$ ,  $B$ 와 세 개의 데이터 세트인  $tra$ ,  $cv$ ,  $wdep$ 를 사용하였다. 인식기의 기각 결과는 곱 근사를 구하는데 포함되었다.

표 2에서 보여진 대로 6개의 인식기 중에서 3개를 선택하여 만들 수 있는 인식기 그룹은 3G1부터 3G20까지 모두 20개이며, 4개를 선택하는 경우에는 4G1부터 4G15까지 모두 15개, 5개를 선택하는 경우에는 5G1부터 5G6까지 모두 6개가 있다. 이들 각 인식기 그룹은 두 테스트 데이터 세트에 관해서 다음의 결합 방법으로 평가되었다. 즉, 표 3에서와 같은 다양한 베이저안 결합 방법과 BKS 방법이 사용되었다. 그림 2,4,6들은  $T$  데이터 세트에 대하여 각각 3,4,5개 인식기 그룹에 대한 실험 결과를 보여주고, 그림 3,5,7들은  $windep$  데이터 세트에 대한 실험 결과를 보여준다.

의존관계의 차수에 대해 2개의 결합 방법이 존재하므로 그림 2~7까지의 결과를 요약하여, 전체 인식기 그룹에 대해서 각 의존관계의 차수에 따라 높은 인식률을

표 1 개별 인식기의 소개

인식기	구조	인식 유형	거리 함수	참고문헌
E1	단일식(singular)	신경망	픽셀 거리 함수	[9]
E2	모듈식(modular)	신경망	방향성 거리 분포	[9]
E3	단일식(singular)	신경망	메쉬 특징	[9]
E4	모듈식(modular)	규칙 기반	수정된 구조적 지식	[10]
E5	모듈식(modular)	규칙 기반	구조적 지식	[10]
E6	단일식(singular)	신경망	칸투어 특징	[8]

표 2 인식기 그룹의 소개

이름	인식기	이름	인식기	이름	인식기
3G1	E1,E2,E3	3G15	E2,E4,E6	4G9	E1,E3,E5,E6
3G2	E1,E2,E4	3G16	E2,E5,E6	4G10	E1,E4,E5,E6
3G3	E1,E2,E5	3G17	E3,E4,E5	4G11	E2,E3,E4,E5
3G4	E1,E2,E6	3G18	E3,E4,E6	4G12	E2,E3,E4,E6
3G5	E1,E3,E4	3G19	E3,E5,E6	4G13	E2,E3,E5,E6
3G6	E1,E3,E5	3G20	E4,E5,E6	4G14	E2,E4,E5,E6
3G7	E1,E3,E6	4G1	E1,E2,E3,E4	4G15	E3,E4,E5,E6
3G8	E1,E4,E5	4G2	E1,E2,E3,E5	5G1	E1,E3,E4,E5,E6
3G9	E1,E4,E6	4G3	E1,E2,E3,E6	5G2	E1,E2,E4,E5,E6
3G10	E1,E5,E6	4G4	E1,E2,E4,E5	5G3	E1,E2,E3,E4,E6
3G11	E2,E3,E4	4G5	E1,E2,E4,E6	5G4	E1,E2,E3,E5,E6
3G12	E2,E3,E5	4G6	E1,E2,E5,E6	5G5	E1,E2,E3,E4,E5
3G13	E2,E3,E6	4G7	E1,E3,E4,E5	5G6	E2,E3,E4,E5,E6
3G14	E2,E4,E5	4G8	E1,E3,E4,E6		

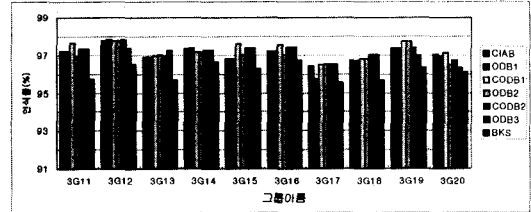
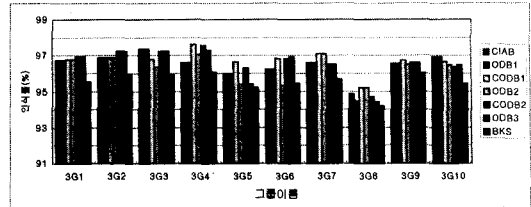


그림 3 windep에 대한 3개 인식기 그룹의 결과

표 3 베이지안 결합 방법의 소개

이름	의미	의존관계
CIAB	조건부 독립 가정에 기반한 베이지안 방법	1차
ODB1	1차 의존관계에 기반한 베이지안 방법	1차
CODB1	조건부 1차 의존관계에 기반한 베이지안 방법	2차
ODB2	2차 의존관계에 기반한 베이지안 방법	2차
CODB2	조건부 2차 의존관계에 기반한 베이지안 방법	3차
ODB3	3차 의존관계에 기반한 베이지안 방법	3차

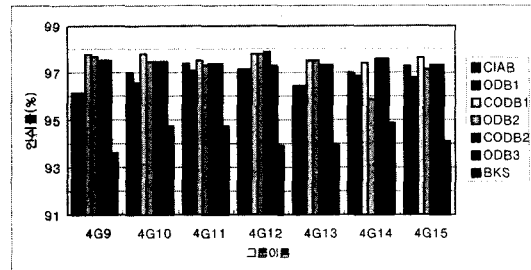
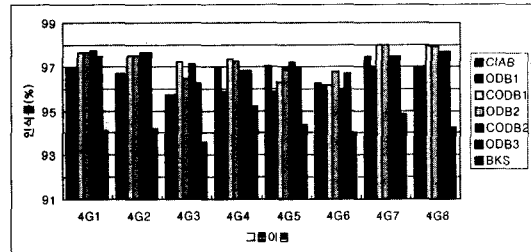


그림 4 T에 대한 4개 인식기 그룹의 결과

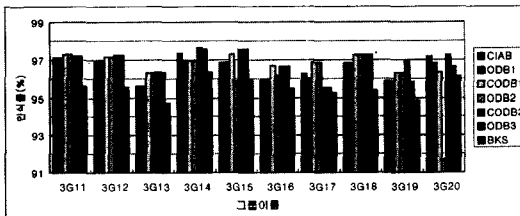
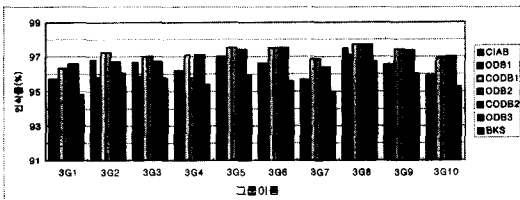


그림 2 T에 대한 3개 인식기 그룹의 결과

보인 결합 방법과 낮은 인식률을 보인 결합 방법끼리 각각 우열을 비교하여 이에 해당하는 인식기 그룹을 작성하면 표 4와 같다. 대부분의 결합 인식률이 95~97% 이므로, 우열을 비교하기 위한 기준은 인식률간의 차가 0.1% 이상인 것으로 정했다. 즉, 두 결합 방법의 인식률간의 차가 0.1% 이내이면, 우열이 없이 동등한 수준(‘=’로 표기)이라고 결정했다. 표 4로부터, 2차나 3차 의존관계는 1차 보다 대부분의 인식기 그룹에서 우수함을

쉽게 알 수 있다. 다만, 2차와 3차 의존관계의 비교에서 보면, 2차가 상대적으로 3차 보다 우수한 경우가 좀 더 많음을 보여주고, 3차는 주로 3개 인식기 그룹에서 우수함을 보여준다.

또한, 전체 인식기 그룹에 대해서 실험에 사용된 의존관계의 차수에 따른 베이지안 결합 방법들간의 성능 비교가 통계적으로 유의한지 알아보기 위하여, 유의수준 0.05(5%)에서 단측검정을 사용하는 t-검정의 쌍체 비교 테스트를 엑셀 프로그램에서 수행하여 판단한 결과가 표 5에 작성되었다. 표의 대각선을 중심으로 좌측 하단은 t-검정을 위해 유의수준과 비교할 수 있는 p-값을 나타내고, 우측 상단은 우열에 따른 비교로 어느 쪽이 통계적으로 유의한지를 ‘>’ 또는 ‘<’과 함께 나타낸다.

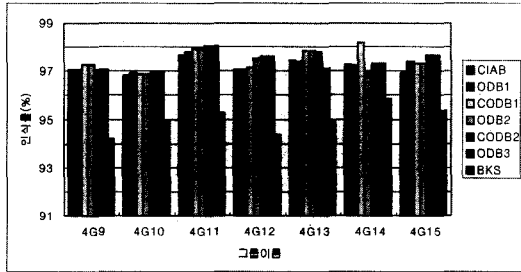
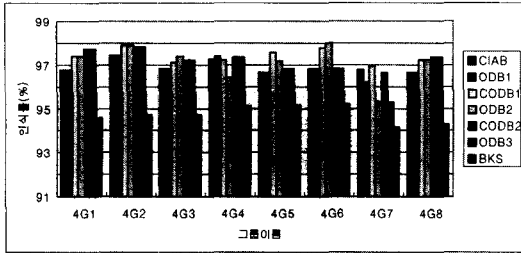


그림 5 *windep*에 대한 4개 인식기 그룹의 결과

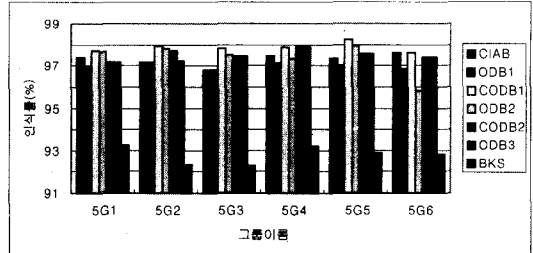


그림 6 *T*에 대한 5개 인식기 그룹의 결과

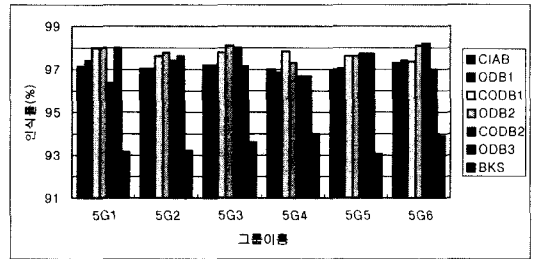


그림 7 *windep*에 대한 5개 인식기 그룹의 결과

빈 칸은 유의한 차이가 없음을 의미한다. *T*에 대한 검정 결과를 보면, 3차 의존관계는 1차 보다 2차에 비해서 더 우수함을 보여주지만, 2차 의존관계는 3차 보다 일부 인식기 그룹에서 우수하나 대체로 비슷한 성능을 보여준다. 또한, *windep*에 대한 검정 결과를 보면, 2차 의존관계는 1차 보다 3차에 비해서 더 우수함을 보여주고, 그러나 3차에 대해서는 혼재된 성능을 보이거나 역시 대체로 비슷한 성능을 보여준다. 즉, 3차 의존관계가 2차

보다 더 우수한 인식기 그룹이 있다.

이러한 실험 결과로부터, 의존관계의 차수가 증가하면서, 2차 또는 3차 의존관계는 거의 1차 의존관계 보다 더 높은 성능을 보여 주며, 3차 의존관계는 5개 인식기로 구성된 그룹을 제외하고는 대체로 2차 의존관계와 비슷한 성능을 보여줌을 알았으며, 때로는 2차 보다 우

표 4 의존관계의 차수에 따른 결합 방법들의 인식률간의 우열 비교

	1차 > 2차	1차 = 2차	1차 < 2차	1차 > 3차	1차 = 3차	1차 < 3차	2차 > 3차	2차 = 3차	2차 < 3차
<i>T</i> 에 대한 높은 인식률간의 비교	3	1	37	4	5	32	20	9	12
- 3개 인식기 그룹	2	0	18	1	2	17	6	7	7
- 4개 인식기 그룹	1	0	14	1	3	11	9	2	4
- 5개 인식기 그룹	0	1	5	2	0	4	5	0	1
<i>T</i> 에 대한 낮은 인식률간의 비교	5	2	34	3	2	36	19	8	14
- 3개 인식기 그룹	3	1	16	2	1	17	7	5	8
- 4개 인식기 그룹	1	1	13	1	0	14	9	2	4
- 5개 인식기 그룹	1	0	5	0	1	5	3	1	2
<i>windep</i> 에 대한 높은 인식률간의 비교	4	7	30	8	9	24	19	8	14
- 3개 인식기 그룹	3	5	12	6	4	10	10	3	7
- 4개 인식기 그룹	1	2	12	1	5	9	6	4	5
- 5개 인식기 그룹	0	0	6	1	0	5	3	1	2
<i>windep</i> 에 대한 낮은 인식률간의 비교	11	6	24	12	8	21	14	7	20
- 3개 인식기 그룹	8	4	8	7	4	9	5	4	11
- 4개 인식기 그룹	3	1	11	2	3	10	4	3	8
- 5개 인식기 그룹	0	1	5	3	1	2	5	0	1

표 5 베이지안 결합 방법들의 t-검정 결과 (유의수준 5%)

T에 대한 t-검정 결과			1차		2차		3차		평균 인식률(%)
			CIAB	ODB1	CODB1	ODB2	CODB2	ODB3	
1차	CIAB	3개 인식기 그룹		> 유의	< 유의		< 유의	< 유의	96.525
		4개 인식기 그룹		> 유의	< 유의	< 유의	< 유의	< 유의	96.830
		5개 인식기 그룹		> 유의	< 유의				97.300
	ODB1	3개 인식기 그룹	0.013		< 유의		< 유의	< 유의	96.368
		4개 인식기 그룹	0.009		< 유의	< 유의	< 유의	< 유의	96.550
		5개 인식기 그룹	0.021		< 유의		< 유의	< 유의	96.983
2차	CODB1	3개 인식기 그룹	0.000	0.000					97.008
		4개 인식기 그룹	0.001	0.000					97.430
		5개 인식기 그룹	0.008	0.000			> 유의	> 유의	97.858
	ODB2	3개 인식기 그룹	0.394	0.220	0.060				97.008
		4개 인식기 그룹	0.008	0.000	0.134				97.277
		5개 인식기 그룹	0.450	0.130	0.052				97.350
3차	CODB2	3개 인식기 그룹	0.000	0.000	0.481	0.102			97.013
		4개 인식기 그룹	0.001	0.000	0.146	0.362			97.330
		5개 인식기 그룹	0.099	0.001	0.018	0.293			97.533
	ODB3	3개 인식기 그룹	0.001	0.000	0.178	0.142	0.079		96.923
		4개 인식기 그룹	0.001	0.000	0.054	0.406	0.170		97.243
		5개 인식기 그룹	0.168	0.006	0.013	0.380	0.182		97.458
windep에 대한 t-검정 결과			1차		2차		3차		평균 인식률(%)
			CIAB	ODB1	CODB1	ODB2	CODB2	ODB3	
1차	CIAB	3개 인식기 그룹			< 유의				96.778
		4개 인식기 그룹			< 유의		< 유의		97.031
		5개 인식기 그룹			< 유의				97.123
	ODB1	3개 인식기 그룹	0.099		< 유의		< 유의		96.726
		4개 인식기 그룹	0.443		< 유의		< 유의		97.039
		5개 인식기 그룹	0.226		< 유의	< 유의			97.172
2차	CODB1	3개 인식기 그룹	0.008	0.003		> 유의		> 유의	97.006
		4개 인식기 그룹	0.000	0.000					97.443
		5개 인식기 그룹	0.002	0.006					97.700
	ODB2	3개 인식기 그룹	0.158	0.339	0.002		< 유의		96.680
		4개 인식기 그룹	0.118	0.100	0.089				97.236
		5개 인식기 그룹	0.000	0.000	0.249			> 유의	97.822
3차	CODB2	3개 인식기 그룹	0.054	0.018	0.064	0.026		> 유의	96.906
		4개 인식기 그룹	0.001	0.000	0.192	0.233			97.344
		5개 인식기 그룹	0.181	0.235	0.234	0.090			97.403
	ODB3	3개 인식기 그룹	0.433	0.246	0.025	0.176	0.043		96.794
		4개 인식기 그룹	0.127	0.080	0.076	0.411	0.096		97.207
		5개 인식기 그룹	0.169	0.190	0.073	0.040	0.456		97.357

수한 예가 있음도 알았다. 더 높은 차수의 의존관계가 더 낮은 차수의 의존관계 보다 높은 성능을 보여 주었다 하더라도, 고차 의존관계가 저차 의존관계 보다 더 높은 성능을 보인다고는 반드시 보장할 수 없다. 이것은 대표적이며 충분한 데이터가 가용하지 않는데 이유가 있다고 여겨질 수 있다. 그렇지만, 의존관계에 기반한 베이지안 결합 방법은 임의의 근사를 취하지 않는 BKS 방법 보다 더 높은 성능을 보여 주었다.

### 5. 결론

이상과 같이, 대부분의 인식기 그룹에서 3차 의존관계가 2차 의존관계 보다 더 높은 성능을 항상 보여 주지는 않는다 하더라도, 3개 인식기 그룹에서 2차 의존관계 보다 높은 성능을 보여 주었다. 3차 의존관계에 기반하여 다수 인식기를 결합하는 방법이 2차 의존관계 보다 더 복잡한 자원을 요구한다 하더라도, 이는 2차 의존관계에 기반한 다수 인식기 결합 방법의 자연스런 확장이다



다. 이상과 같이, 본 논문은 3차 의존관계가 2차 의존관계 보다 무조건 좋다는 점을 강조하려는 것이 아니고, 2차 의존관계에 의한 결합 방법을 3차 의존관계로까지 확장하기 위한 이론적 배경과 관련 알고리즘에 대해서 제안하였으며, 이를 Concordia 대학과 UCI 대학의 필기 숫자 데이터를 사용하여 실험하였다. 의존관계의 차수를 늘리는 것이 근사에 따른 오차를 줄여서 어느 정도 인식율의 향상에 기여할 수 있지만, 최적의 곱 근사를 구하기 위한 복잡도가 증가하고, 지나친 적합이나 데이터의 부족을 겪을 수 있는 위험이 있다. 게다가 3차 의존관계를 고려할 때, 훈련 데이터의 최적의 곱 근사가 테스트 데이터의 것과 같은 데이터 세트를 구하는 것이 실제 응용에서 종종 어렵다. 따라서, 충분하고 대표성을 지닌 샘플 데이터 세트를 확보하거나, 최상의 성능을 위한 의존관계의 차수를 적절히 선정하거나, 고차 의존관계의 고려에 따른 복잡도를 줄이기 위한 연구를 다루는 것이 유용하리라 여긴다. 그 외에, 이론적 근거에서 성능 향상을 알아보기 위한 의존관계에 기반한 베이저안 방법의 하위 경계와 상위 경계를 연구하는 것이 유용하다고 본다.

### 참 고 문 헌

- [1] L. Xu and A. Krzyzak, and C. Y. Suen, "Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition," IEEE Trans. on Systems, Man and Cybernetics, vol.22, no.3, pp.418-435, 1992.
- [2] 강희중, 이성환, "무제약 필기 숫자를 인식하기 위한 다수 인식기를 결합하는 의존관계 기반의 프레임워크", 정보과학회논문지 : 소프트웨어 및 응용, 제27권, 제8호, pp.855-863, 2000.
- [3] Y. S. Huang and C. Y. Suen, "A Method of Combining Multiple Experts for the Recognition of Unconstrained Handwritten Numerals," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.17, no.1, pp.90-94, 1995.
- [4] Mandler, E. and Schuermann, J., "Combining the classification results of independent classifiers based on the dempster/shafer theory of evidence," In E. S. Gelsema and L. N. Kanal, editors, Pattern Recognition and Artificial Intelligence, pp. 381-393. 1988.
- [5] D.-S. Lee and S. N. Srihari, "Handprinted Digit Recognition : A Comparison of Algorithms," Proc. of the 3rd Int. Workshop on Frontiers in Handwriting Recognition, pp.42-45, 1993.
- [6] C. K. Chow and C. N. Liu, "Approximating Discrete Probability Distributions with Dependence Trees," IEEE Trans. on Information Theory, vol. 14, no.3, pp.462-467, 1968.
- [7] P. M. Lewis, "Approximating Probability Distribu-

tions to Reduce Storage Requirement," Information and Control, vol.2, pp.214-225, Sep., 1959.

- [8] Matsui, T., Tsutsumida, T., and Srihari, S. N., "Combination of Stroke/Background Structure and Contour-direction Features in Handprinted Alphabetic Recognition," In Proceedings of the 4th Int. Workshop on Frontiers in Handwriting Recognition, pp. 87-96, 1994.
- [9] Oh, I.-S. and Suen, C. Y., "Distance features for neural network-based recognition of handwritten characters," International Journal on Document Analysis and Recognition, 1(2):73-88, 1998.
- [10] Oh, I.-S., Lee, J.-S., Hong, K.-C., and Choi, S.-M., "Class-expert approach to unconstrained handwritten numeral recognition," In Proceedings of the 5th Int. Workshop on Frontiers in Handwriting Recognition, pp. 35-40, 1996.
- [11] Suen, C. Y., Nadal, C., Legault, R., Mai, T. A., and Lam, L., "Computer Recognition of Unconstrained Handwritten Numerals," In Proceedings of IEEE, pp. 1162-1180, 1992.
- [12] Blake, C. and Merz, C., UCI repository of machine learning databases [http://www.ics.uci.edu/~mllearn/mlrepository.html], Irvine, CA, Dept. of Information and Computer Sciences, 1998.

### 강 희 중

1986년 서울대학교 전자계산기공학과 졸업(학사). 1988년 한국과학기술원 전산학과 졸업(석사). 1997년 한국과학기술원 전산학과 졸업(박사). 1986년~1998년 삼성전자주식회사 기업통신개발그룹 선임 연구원. 1998년~2000년 고려대학교 인공지능연구센터 연구교수. 2000년~현재 한성대학교 컴퓨터공학부 교수. 2002년 8월~현재 미국 University of Maryland, College Park, 방문 연구원. 관심분야는 인공지능, 패턴인식, 그룹웨어 임